Robust Training Sequence Design for Spatially Correlated MIMO Channel Estimation

Chin-Te Chiang and Carrson C. Fung, *Member, IEEE*

***Abstract*—A robust superimposed training sequence design is proposed for spatially correlated multiple-input–multiple-output (MIMO) channel estimation. The proposed scheme does not require accurate knowledge of the spatial correlation matrix, and it is shown to outperform previously proposed robust correlated MIMO channel estimators, such as relaxed minimum mean square error (RMMSE) and least-square RMMSE. Since the training sequence is overlaid into the data stream, the spectral efficiency of the system is higher than those that use time-multiplexed pilots. A solution for the sequence can easily be obtained by using a projection on convex-set-based iterative algorithm that is guaranteed to converge as long as the training sequence matrix is initialized to have full rank. Furthermore, it is shown that the proposed scheme is identical to the RMMSE-based schemes when the MIMO channel is spatially uncorrelated. The computational complexity of the proposed algorithm is also illustrated.**

***Index Terms*—Affine precoder, majorization, multiple-input–multiple-output (MIMO), robust channel estimation, spatial correlation, superimposed training (SIT) sequence.**

I. Introduction

THE DEVELOPMENT of single-user multiple-input–multiple-output (MIMO) systems [1]–[3] has spurred tremendous research effort in advancing techniques that maximize diversity and spatial multiplexing gains [4], [5]. To realize such gains, channel state information (CSI) must accurately be obtained. Although techniques such as differential space–time coding [6], [7] and differential orthogonal space–time block coding (STBC) [8], [9] have been proposed to blindly demodulate and decode the received signal, this degrades both performance (compared with coherent techniques) and spectrum efficiency. Therefore, coherent detection is widely used in current MIMO systems where CSI is usually obtained using time-multiplexed pilot symbols.

Theoretically, MIMO performance gain depends only on the minimum between the number of transmit and receive antennas. However, propagation limitations such as channel rank loss and antenna correlation have to be properly handled before such performance gains can be realized. Increased antenna correlation can be attributed to a reduction in antenna spacing or angular spread, which is caused by the lack of a rich scattering environment around the transceiver. Hence, some degree of spatial correlation will be experienced at the transmitter and/or receiver. Hence, it is vital to account for spatial correlation during the channel estimation process to maximize the decoding performance.

Many techniques have been proposed to tackle the problem of correlated MIMO channel estimation. Reference [10] proposed using a state-space approach to estimate and track time-varying correlated MIMO channels, where the channel correlation matrix is estimated from the received data and treated as part of the state variable. In [11], a precoder-assisted linear minimum mean-square error (MMSE) estimator was proposed to estimate the channel. In [12], two channel estimators were derived under the MMSE and conditional mutual information criteria by exploiting the virtual channel representation. Unfortunately, there is no closed-form solution; thus, the solution has to be computed numerically. Another MMSE-based channel estimator was derived in [13] using structured correlation, which allows it to obtain better mean-square error (MSE) performance than the unstructured-based MMSE estimator. One major drawback shared among these estimators is that they require exact knowledge about the spatial correlation to outperform channel estimators that take no such correlation into account. Another disadvantage is that they were all derived under the premise that time-multiplexed pilot symbols are used, which can drastically reduce the transmission efficiency, particularly in cases where the channel is undergoing fast fading.

To bypass the second problem, a superimposed training (SIT) sequence-based channel estimation algorithm was proposed in [14], where the SIT sequence is arithmetically added into the transmitted signal, thus allowing the system to free up valuable time slots that were previously used by time-multiplexed pilot symbols. The training sequence can also be used to deal with the problem of synchronization [17]. Improved channel estimation algorithms based on the SIT sequence have since appeared in the literature [15]–[17]. The sequence itself can be extracted at the receiver by using first-order statistics [15], [16] or by using affine precoding [18]–[20]. However, the effectiveness of these algorithms still hinges on acquiring accurate estimates of the spatial correlation, making these methods somewhat infeasible in real situations. To combat against this problem, the relaxed MMSE (RMMSE) and least-square RMMSE (LS-RMMSE) algorithms that have recently been proposed by [21] can circumvent the dependency on the correlation matrix by

Manuscript received July 13, 2010; revised April 29, 2011; accepted June 4, 2011. Date of publication July 14, 2011; date of current version September 19, 2011. This work was supported by the National Science Council under Grant 99-2219-E-009-011. This paper was presented in part at the 2010 IEEE International Conference on Communications. The review of this paper was coordinated by Dr. C. Cozzo.

C.-T. Chiang was with the Department of Electronics Engineering, National Chiao Tung University, Hsinchu 300, Taiwan. He is now with the Intellectual Property Office, Ministry of Economic Affairs, Taipei 106, Taiwan (e-mail: jeremiah1214@gmail.com).

C. C. Fung is with the Department of Electronics Engineering, National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: c.fung@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2011.2161640

using diagonal loading. A different approach using basis expansion was proposed in [22].

In this paper, a deterministic approach is taken in the design of SIT sequences that are robust toward spatial correlation uncertainty. Such a training sequence is then applied to the MMSE estimator to estimate the channel. The approach taken in this work allows for accurate and robust channel estimation without *a priori* knowledge about the channel distribution model. For ease of presentation, such an estimator will be called RoMMSE estimator (despite the fact that it is the sequence that is robust to spatial correlation error). The proposed design exploits the affine precoder scheme proposed in [20] and [24] to extract the training sequence for channel estimation. Since the training sequence is overlaid onto the information-bearing signal, it allows for greater spectral efficiency compared with conventional systems that use time-multiplexed pilot symbols. The simulation results will show that the proposed scheme performs extremely well against the estimator in [20], which does not take into account the spatial correlation estimate error. Moreover, the RoMMSE estimator also outperforms the RMMSE and LS-RMMSE estimators when the MIMO channels are spatially correlated. Finally, the RoMMSE estimator will also be compared, analytically and by simulation, to the RMMSE and LS-RMMSE estimators for uncorrelated MIMO channels in which it is shown that the three estimators are identical. The rest of this paper is organized as follows: The system model and the affine precoding method are given in Section II, followed by a detailed description of the proposed RoMMSE algorithm in Section III. Simulation results are provided in Section IV, and this paper is concluded in Section V.

Notation: Upper (lower) boldface letters indicate matrices (column vectors). Superscript ^{H} denotes Hermitian, ^{T} denotes transposition, and * denotes conjugation. $E[\cdot]$ stands for expectation. $\mathcal{N}(A)$ denotes the null space of $\mathbf{A}$. The operation $\mathrm{vec}(\mathbf{A})$ forms a column vector by vertically stacking the column vectors of $\mathbf{A}$. $\mathrm{tr}(\mathbf{A})$ denotes the trace of the matrix $\mathbf{A}$. $\mathrm{diag}(x)$ denotes a diagonal matrix with x on its main diagonal, $\mathbf{I}_N$ denotes an $N \times N$ identity matrix, and $\mathbf{0}_{M\times N}$ denotes an $M \times N$ all zero matrix. $\otimes$ denotes the Kronecker product. $\|\mathbf{A}\|_F$ denotes the Frobenius norm of the matrix $\mathbf{A}$.

II. System Model and Affine Precoding

A. System Model

The system model used in [20] is adopted herein. For the sake of completeness, the model will also be described in the sequel. Consider a spatially correlated flat-fading MIMO channel with N_t transmit and N_r receive antennas, as shown in Fig. 1. The information-bearing signal vector is denoted as $\mathbf{u}(k) = [u(kN_s)\, u(kN_s+1) \cdots u(kN_s+N_s-1)]^T$, where k is the block index, and N_s denotes the block size. Each block of the signal is encoded using STBC, which can be used to increase the transmit diversity or multiplexing gain [23]. The STBC has N_t number of output vectors, with each vector containing $K \geq N_t$ symbols as full rate STBC is assumed. This can be represented in matrix form as $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2 \cdots \mathbf{x}_{N_t}]^T \in \mathbb{C}^{N_t \times K}$, where $\mathbf{x}_i \in \mathbb{C}^K$, $i = 1, 2, \ldots, N_t$ denotes the ith output vector. Each vector is then fed into the precoder $\mathbf{P} = [\mathbf{p}_1\ \mathbf{p}_2 \cdots \mathbf{p}_K]^T \in \mathbb{C}^{K\times(K+L)}$, which adds $L \geq N_t$ redundant symbols to each block of signal, resulting in the output signal vector $\mathbf{d}_i \in \mathbb{C}^{K+L}$ for $i = 1, 2, \ldots, N_t$. All N_t outputs of the precoder can be represented in matrix form as

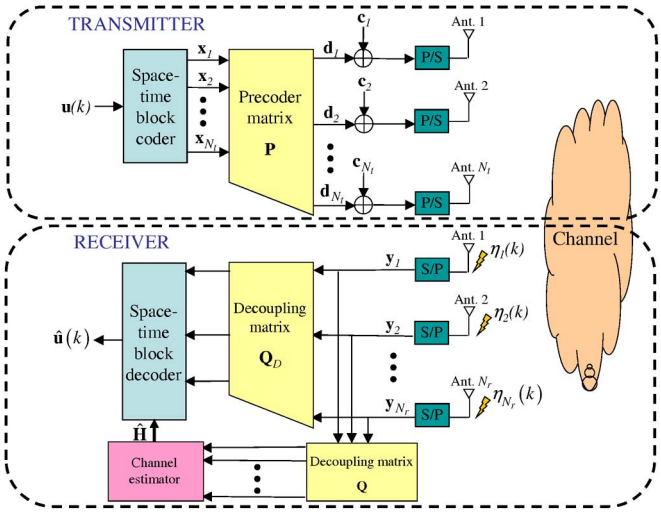

Fig. 1. Block diagram of MIMO transceiver.

$$\mathbf{D} \triangleq \begin{bmatrix} \mathbf{d}_1^T \\ \mathbf{d}_2^T \\ \vdots \\ \mathbf{d}_{N_t}^T \end{bmatrix} = \mathbf{XP} = \begin{bmatrix} \mathbf{x}_1^T\mathbf{P} \\ \mathbf{x}_2^T\mathbf{P} \\ \vdots \\ \mathbf{x}_{N_t}^T\mathbf{P} \end{bmatrix} \in \mathbb{C}^{N_t\times(K+L)}. \quad (1)$$

As seen in the sequel, the precoder is used to assist in the channel estimation [8]–[10] by eliminating the information-bearing signal at the receiver, thus leaving the SIT sequence intact for channel estimation. It was shown in [24] that the precoder can also be designed to improve the symbol detection rate or minimize the mean square error between the transmitted and recovered signals [8]. After precoding, the SIT sequence vector $\mathbf{c}_i$, $i = 1, 2, \ldots, N_t$ is added to $\mathbf{d}_i$. Each vector is then serialized before it is transmitted across the flat-fading MIMO channel, which is represented in matrix form as $\mathbf{H} \in \mathbb{C}^{N_r\times N_t}$. Thus, the received signal can be written as

$$\mathbf{Y} = \mathbf{H}(\mathbf{C}+\mathbf{D}) + \boldsymbol{\eta} = \mathbf{HC} + \mathbf{HXP} + \boldsymbol{\eta} \quad (2)$$

where

$$\mathbf{C} \triangleq \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_{N_t}^T \end{bmatrix} \in \mathbb{C}^{N_t\times(K+L)}, \quad \text{and } \boldsymbol{\eta} \in \mathbb{C}^{N_r\times(K+L)}$$

are the SIT sequence matrix and the additive channel noise matrix, respectively. Notice in (2) that the received signal in space lies in the rows of $\mathbf{Y}$. Thus, the rows of the information-bearing portion of the signal, i.e., $\mathbf{x}_i^T\mathbf{P}$, $i = 1, 2, \ldots, N_t$, belong to the row space of $\mathbf{P}$. Hence, the rows of $\mathbf{HXP}$ also belong to the same subspace. This is different from the conventional model used in [9], [25], and [26], where the information-bearing portion of the received signal is embedded inside the

range space of the unknown channel matrix, thus making it difficult for channel estimation using the SIT sequence. The affine precoding approach adopted herein eases the decoupling of the information-bearing signal and the training sequence because decoupling can now be done by postmultiplying $\mathbf{Y}$ by a decoupling matrix $\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2 \dots \mathbf{q}_{K+L}]^T \in \mathbb{C}^{(K+L)\times N_t}$, resulting in

$$\mathbf{YQ} = \mathbf{HCQ} + \mathbf{HXPQ} + \boldsymbol{\eta}\mathbf{Q}. \tag{3}$$

Thus, by requiring the columns of $\mathbf{Q}$ to lie in $\mathcal{N}(\mathbf{P})$, i.e., $\mathbf{PQ} = \mathbf{0}_{K\times N_t}$, then (3) becomes

$$\mathbf{YQ} = \mathbf{HCQ} + \boldsymbol{\eta}\mathbf{Q}. \tag{4}$$

In other words, the training sequence vector $\mathbf{c}_i, \ i = 1, 2, \dots, N_t$, should lie in the column space of $\mathbf{Q}$. Therefore, the condition that $\mathbf{CP}^H = \mathbf{0}_{N_t\times K}$ guarantees the subspaces spanned by the vectors in $\mathbf{P}$ and $\mathbf{C}$ are complementary [20]. This suggests a simple way to design $\mathbf{P}$ and $\mathbf{Q}$ by extracting components off of an orthogonal matrix, i.e.,

$$\mathbf{P} = \sqrt{\frac{K+L}{K}}\mathbf{O}(1:K,:) \in \mathbb{C}^{K\times(K+L)}$$
$$\mathbf{Q} = \left(\mathbf{O}\left((K+1):(K+N_t),:\right)\right)^H \in \mathbb{C}^{(K+L)\times N_t}.$$

Note that $\mathbf{O}(1:K,:)$ and $\mathbf{O}((K+1):(K+N_t),:)$ keep only rows 1 to K and rows $(K+1)$ to $(K+N_t)$ of an orthogonal matrix $\mathbf{O} \in \mathbb{C}^{(K+N_t)\times(K+L)}$ [24]. Hence, $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}_{N_t}$ so that noise amplification will not occur in the channel estimation process.

In addition to channel estimation, another decoupling matrix $\mathbf{Q}_D$ can be designed to maximize the symbol detection performance. Such a decoupling matrix can be chosen to satisfy the condition $\mathbf{Q}_D = \mathbf{P}^H(\mathbf{PP}^H)^{-1}$, where $\mathbf{P}$ is designed such that $\mathbf{CP}^H = \mathbf{0}_{N_t\times K}$. This ensures that the detection process is free of interference from the SIT sequence when $\mathbf{Q}_D$ is postmultiplied to $\mathbf{Y}$. Therefore

$$\mathbf{PP}^H = \left(\mathbf{Q}_D^H\mathbf{Q}_D\right)^{-1} = \frac{K+L}{K}\mathbf{I}_K$$

such that $\mathrm{tr}(\mathbf{PP}^H) = K + L$. This is to ensure that the average transmitted power of the information-bearing signal is unchanged after precoding.

According to the Kronecker model [23], the channel matrix can be decomposed as

$$\mathbf{H} = \boldsymbol{\Sigma}_r^{\frac{1}{2}}\mathbf{H}_w\boldsymbol{\Sigma}_t^{\frac{1}{2}} \tag{5}$$

where $\boldsymbol{\Sigma}_r^{1/2} \in \mathbb{C}^{N_r\times N_r}$ and $\boldsymbol{\Sigma}_t^{1/2} \in \mathbb{C}^{N_t\times N_t}$ are the Cholesky factors of the spatial correlation matrix of the receiver and transmitter, respectively. Hence, the overall spatial correlation is $\mathbf{R} = \boldsymbol{\Sigma}_t \otimes \boldsymbol{\Sigma}_r$. The entries of $\mathbf{H}_w \in \mathbb{C}^{N_r\times N_t}$ are independent and identically distributed zero-mean complex Gaussian random variables with unit variance. Thus, $E[\mathrm{vec}(\mathbf{H}_w)\mathrm{vec}^H(\mathbf{H}_w)] = \mathbf{I}_{N_rN_t}$.

B. MMSE Estimator and Training Sequence Design

To derive the proposed RoMMSE estimator, (4) is first vectorized to obtain the received signal vector

$$\mathbf{y} = \widetilde{\mathbf{C}}\mathbf{h} + \mathbf{n} \tag{6}$$

where $\mathbf{y} = \mathrm{vec}(\mathbf{YQ}) \in \mathbb{C}^{N_rN_t}$, $\widetilde{\mathbf{C}} = (\mathbf{CQ})^T \otimes \mathbf{I}_{N_r} \in \mathbb{C}^{N_rN_t\times N_rN_t}$, $\mathbf{h} = \mathrm{vec}(\mathbf{H}) \in \mathbb{C}^{N_rN_t}$, and $\mathbf{n} = \mathrm{vec}(\boldsymbol{\eta}\mathbf{Q}) \in \mathbb{C}^{N_rN_t}$. $E[\mathbf{nn}^H] = \sigma_n^2\mathbf{I}_{N_rN_t}$. From the vectorized received signal $\mathbf{y}$, the linear MMSE estimator of $\mathbf{h}$ is [27, p. 387]

$$\widehat{\mathbf{h}} = \mathbf{R}_{\mathbf{yh}}^H\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{y} = \mathbf{R}\widetilde{\mathbf{C}}^H\left(\widetilde{\mathbf{C}}\mathbf{R}\widetilde{\mathbf{C}}^H + \sigma_{\mathbf{nn}}^2\mathbf{I}_{N_rN_t}\right)^{-1}\mathbf{y} \tag{7}$$

where $\mathbf{R}_{\mathbf{yy}} \triangleq E[\mathbf{yy}^H]$, $\mathbf{R}_{\mathbf{yh}} \triangleq E[\mathbf{yh}^H]$, and $\mathbf{R} \triangleq E[\mathbf{hh}^H]$ are the autocorrelation matrix of the received signal $\mathbf{y}$, the cross-correlation matrix of $\mathbf{y}$ and $\mathbf{h}$, and the spatial correlation matrix of the channel, respectively. All the matrices are of size $N_rN_t \times N_rN_t$. Therefore, the optimal MMSE estimate of $\mathbf{h}$ can be obtained by finding the optimal training sequence matrix $\widetilde{\mathbf{C}}$. Note that the mean-square error matrix between $\mathbf{h}$ and $\widehat{\mathbf{h}}$ is written as [27, p. 387]

$$\begin{aligned}\boldsymbol{\xi} &= E\left[(\mathbf{h}-\widehat{\mathbf{h}})(\mathbf{h}-\widehat{\mathbf{h}})^H\right] \\ &= \left(\mathbf{R}^{-1} + \widetilde{\mathbf{C}}^H\left(\sigma_{\mathbf{nn}}^2\mathbf{I}_{N_rN_t}\right)^{-1}\widetilde{\mathbf{C}}\right)^{-1}.\end{aligned} \tag{8}$$

From (8), [20] proposed to design the optimal training sequence matrix $\widetilde{\mathbf{C}}$ by minimizing the trace of $\boldsymbol{\xi}$ subject to the power constraint $\mathrm{tr}(\mathbf{CC}^H) \le N_t(K+L)\sigma_{\mathbf{cc}}^2 \triangleq P_T$, where $\sigma_{\mathbf{cc}}^2$ is the average power of the training sequence. It was assumed in [20] that the average transmitted power, which includes the power of the information-bearing and training signals, is normalized as $\sigma_{\mathbf{xx}}^2 + \sigma_{\mathbf{cc}}^2 = 1$, where $\sigma_{\mathbf{xx}}^2$ is the variance of the information-bearing signal. This assumption will also be applied to the proposed RoMMSE estimator. Since $\widetilde{\mathbf{C}} = (\mathbf{CQ})^T \otimes \mathbf{I}_{N_r}$, using the properties that $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$, $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$, $(\mathbf{A}\otimes\mathbf{B})^H = \mathbf{A}^H\otimes\mathbf{B}^H$, and $\mathrm{tr}(\mathbf{A}\otimes\mathbf{B}) = \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})$, the power constraint on $\widetilde{\mathbf{C}}$ can be written as

$$\begin{aligned}\mathrm{tr}(\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}^H) &= \mathrm{tr}\left(\left[(\mathbf{CQ})^T\otimes\mathbf{I}_{N_r}\right]\left[(\mathbf{CQ})^T\otimes\mathbf{I}_{N_r}\right]^H\right) \\ &= \mathrm{tr}\left(\left[(\mathbf{CQ})^T\otimes\mathbf{I}_{N_r}\right]\left[(\mathbf{CQ})^*\otimes\mathbf{I}_{N_r}\right]\right) \\ &= \mathrm{tr}\left(\left[(\mathbf{CQ})^T(\mathbf{CQ})^*\right]\otimes\mathbf{I}_{N_r}\right) \\ &= \mathrm{tr}\left((\mathbf{CQ})^T(\mathbf{CQ})^*\right)\mathrm{tr}(\mathbf{I}_{N_r}) \\ &= N_r\mathrm{tr}(\mathbf{Q}^T\mathbf{C}^T\mathbf{C}^*\mathbf{Q}^*) \\ &= N_r\mathrm{tr}(\mathbf{C}^T\mathbf{C}^*) \\ &\le N_rP_T \triangleq \widetilde{P}_T.\end{aligned} \tag{9}$$

The inequality is obtained because $\mathrm{tr}(\mathbf{C}^T\mathbf{C}^*) = \mathrm{tr}(\mathbf{C}^H\mathbf{C}) = \mathrm{tr}(\mathbf{CC}^H) = \|\mathbf{C}\|_F^2 \le P_T$. It is important to note that the performance of the RoMMSE estimator is dependent on the total transmission power P_T and not the number of redundant vectors L. The latter is, however, necessary to allow for the

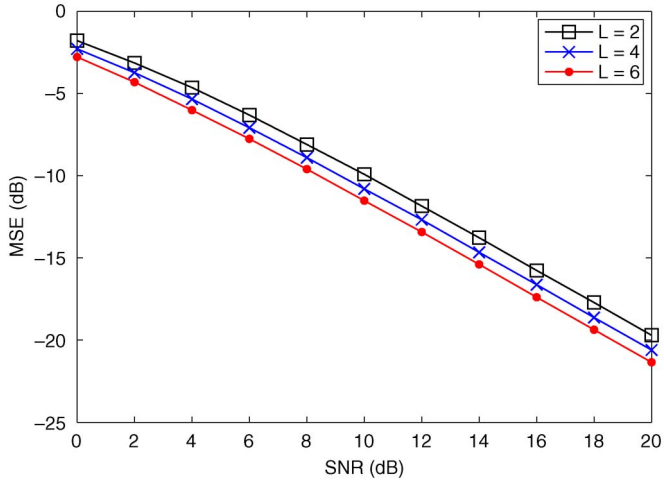

Fig. 2. MSE versus SNR performance comparison between different numbers of redundant vectors with different P_T for spatially correlated 2×2 MIMO system, $\Delta = 5^\circ$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

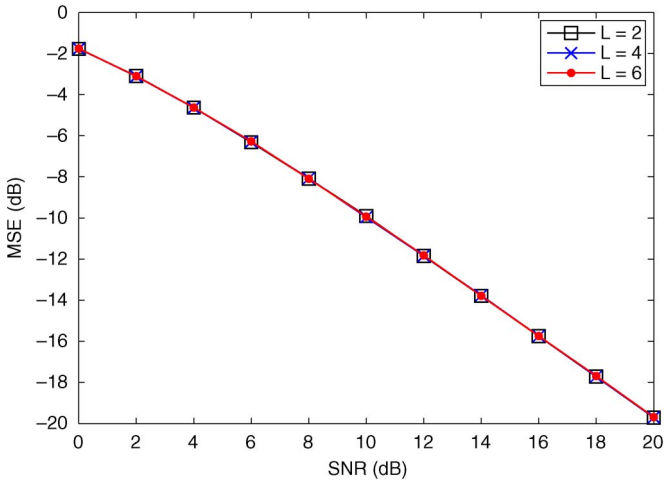

Fig. 3. MSE versus SNR performance comparison between different numbers of redundant vectors with fixed P_T for spatially correlated 2×2 MIMO system, $\Delta = 5^\circ$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

SIT sequence to be decoupled from the information-bearing signal at the receiver. Figs. 2 and 3 show the MSE performance of the proposed RoMMSE estimator (to be described in the next section) for 2×2 correlated MIMO channels when L is increased with different and fixed P_T, respectively. In the latter case, $\sigma^2_{\mathbf{cc}}$ is decreased while L is increased to keep P_T constant. From Fig. 2, it is clear that the MSE performance improves as L increases, whereas Fig. 3 shows that such performance improvement is due to the increase in P_T, not just L, because as P_T is kept constant even while L is increased, there is no change in MSE performance.

III. Proposed RoMMSE Design

A. Proposed Training Sequence Design

It is clear from (8) that exact knowledge of $\mathbf{R}$ is required at the receiver to obtain an accurate estimate of $\mathbf{h}$ using (7). However, in all likelihood, only an estimate of $\mathbf{R}$ can be obtained, for example, using the method proposed in [28]. To desensitize the MSE from the estimation error of $\mathbf{R}$, a novel SIT sequence design is proposed herein to incorporate such an estimation error. As the spatial correlation matrix is estimated at the receiver before it is fed back to the transmitter using a low-rate control channel, the quantization error will tend to color the spatial correlation mismatch between the estimated and actual spatial correlations. Hence, a deterministic approach is proposed herein to bound the error in a norm ball. Applying such a SIT sequence into the MMSE estimator in (7) allows the estimator to be more robust against estimation error in the spatial correlation than other MMSE-based estimators that do not take such error into account. Although the rate of change of the channel statistics is slower than that of the channel coefficients, imperfect channel statistics will still adversely affect the channel estimation performance and thus the bit error rate (BER), if not properly accounted for in the system design. Moreover, better robust channel estimation can be obtained if the spatial correlation and channel coefficient mismatches can separately be accounted for as the structure of the spatial correlation mismatch matrix will be different from that of the channel coefficient matrix.

Let

$$\mathbf{R} = \widehat{\mathbf{R}} + \mathbf{E} \tag{10}$$

where $\widehat{\mathbf{R}}$ denotes the estimate of $\mathbf{R}$, and $\mathbf{E}$ is its corresponding spatial correlation mismatch matrix, respectively. In the present scheme, the error power is upper bounded such that $\|\mathbf{E}\|_F \leq \varepsilon$, where ε is a predefined error power bound. In practice, a table of sequences for different ε can be computed *a priori* and placed in memory at the transmitter. An adaptive approach can then be used to choose a suitable sequence for channel estimation.

Using this bound with (9) and (10), the training sequence matrix $\mathbf{C}$ (or its equivalent $\widetilde{\mathbf{C}}$) can be designed by minimizing the maximum mean square error $\boldsymbol{\xi}$, i.e.,

$$\min_{\|\widetilde{\mathbf{C}}\|_F^2 \leq \widetilde{P}_T} \max_{\|\mathbf{E}\|_F \leq \varepsilon} \mathrm{tr}\left(\left[(\widehat{\mathbf{R}}+\mathbf{E})^{-1} + \widetilde{\mathbf{C}}^H \left(\sigma^2_{\mathbf{nn}} \mathbf{I}_{N_r N_t} \right)^{-1} \widetilde{\mathbf{C}} \right]^{-1} \right). \tag{11}$$

Note that (11) is not a convex problem with respect to $\mathbf{E}$ and $\mathbf{C}$. However, the problem can be decomposed into two separate convex optimization problems: one with respect to $\mathbf{E}$ and the other to $\mathbf{C}$. Furthermore, performing SVD on $\widetilde{\mathbf{C}}$, i.e., $\widetilde{\mathbf{C}} = \mathbf{U}_{\widetilde{\mathbf{C}}} \boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}} \mathbf{V}^H_{\widetilde{\mathbf{C}}}$ and using the property $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$, the objective function of the maximization problem in (11) can be rewritten as

$$\begin{aligned} &\mathrm{tr}\left(\left[(\widehat{\mathbf{R}} + \mathbf{E})^{-1} + \widetilde{\mathbf{C}}^H \left(\sigma^2_{\mathbf{nn}} \mathbf{I}_{N_r N_t} \right)^{-1} \widetilde{\mathbf{C}} \right]^{-1} \right) \\ &\quad = \mathrm{tr}\left(\left[(\widehat{\mathbf{R}} + \mathbf{E})^{-1} + \sigma^{-2}_{\mathbf{nn}} \mathbf{V}_{\widetilde{\mathbf{C}}} \boldsymbol{\Sigma}^H_{\widetilde{\mathbf{C}}} \boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}} \mathbf{V}^H_{\widetilde{\mathbf{C}}} \right]^{-1} \right) \\ &\quad = \mathrm{tr}\left(\left[\mathbf{V}^H_{\widetilde{\mathbf{C}}} (\widehat{\mathbf{R}} + \mathbf{E})^{-1} \mathbf{V}_{\widetilde{\mathbf{C}}} + \sigma^{-2}_{\mathbf{nn}} \boldsymbol{\Sigma}^H_{\widetilde{\mathbf{C}}} \boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}} \right]^{-1} \right). \end{aligned} \tag{12}$$

Next, using the property $\mathrm{tr}(\mathbf{A}+\mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$ and the matrix inversion lemma $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$, and letting $\mathbf{A} = \sigma^{-2}_{\mathbf{nn}} \boldsymbol{\Sigma}^H_{\widetilde{\mathbf{C}}} \boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}$,

$\mathbf{C} = \mathbf{V}_{\widetilde{\mathbf{C}}}^H(\widehat{\mathbf{R}}+\mathbf{E})^{-1}\mathbf{V}_{\widetilde{\mathbf{C}}}$ and $\mathbf{B} = \mathbf{D} = \mathbf{I}_{N_rN_t}$, then (12) becomes

$$\begin{aligned}
&\mathrm{tr}\Bigg(\sigma_{\mathbf{nn}}^2\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-H} - \sigma_{\mathbf{nn}}^2\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-H} \\
&\quad\times\left[\sigma_{\mathbf{nn}}^2\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-H} + \mathbf{V}_{\widetilde{\mathbf{C}}}^H(\widehat{\mathbf{R}}+\mathbf{E})\mathbf{V}_{\widetilde{\mathbf{C}}}\right]^{-1}\sigma_{\mathbf{nn}}^2\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-H}\Bigg) \\
&= \mathrm{tr}\left(\sigma_{\mathbf{nn}}^2\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}\right) \\
&\quad - \sigma_{\mathbf{nn}}^4\mathrm{tr}\left(\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}\left[\sigma_{\mathbf{nn}}^2\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1} + \mathbf{V}_{\widetilde{\mathbf{C}}}^H(\widehat{\mathbf{R}}+\mathbf{E})\mathbf{V}_{\widetilde{\mathbf{C}}}\right]^{-1}\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}\right) \\
&= \mathrm{tr}\left(\sigma_{\mathbf{nn}}^2\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}\right) \\
&\quad - \sigma_{\mathbf{nn}}^2\mathrm{tr}\left(\left[\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}} + \sigma_{\mathbf{nn}}^{-2}\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}\mathbf{V}_{\widetilde{\mathbf{C}}}^H(\widehat{\mathbf{R}}+\mathbf{E})\mathbf{V}_{\widetilde{\mathbf{C}}}\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}\right]^{-1}\right) \\
&= \mathrm{tr}\left(\sigma_{\mathbf{nn}}^2\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}\right) \\
&\quad - \sigma_{\mathbf{nn}}^2\mathrm{tr}\left(\left[\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}} + \sigma_{\mathbf{nn}}^{-2}\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}(\widehat{\mathbf{R}}+\mathbf{E})\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}\right]^{-1}\right)
\end{aligned} \tag{13}$$

where $\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}} \triangleq \boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^H\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}$. Since the first term of (13) does not depend on $\mathbf{E}$, maximizing (11) is equivalent to minimizing the second term in (13). Therefore, the maximization problem in (11) becomes

$$\min_{\|\mathbf{E}\|_F\le\varepsilon}\mathrm{tr}\left(\left[\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}} + \sigma_{\mathbf{nn}}^{-2}\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}(\widehat{\mathbf{R}}+\mathbf{E})\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}\right]^{-1}\right) \tag{14}$$

and it can easily be solved using convex optimization toolbox such as cvx [29] since the $(\widehat{\mathbf{R}}+\mathbf{E})^{-1}$ term is eliminated.

Unfortunately, even if (14) is substituted into (11), it is difficult to find a closed-form solution for $\mathbf{C}$. Therefore, the iterative algorithm in Fig. 4 is proposed. As seen from the figure, $\widetilde{\mathbf{C}}$ is first initialized to be a full rank matrix satisfying the condition $\mathbf{C}(0)\mathbf{P}^H = \mathbf{0}$. $\mathbf{C}(0)$ is then used in (17) [or, equivalently, (14)] to solve for a solution for $\mathbf{E}$. This is then used in (18) to solve for $\widetilde{\mathbf{C}}$. This process will be repeated until $\|\mathbf{E}(n)-\mathbf{E}(n-1)\|^2/\varepsilon$ is less than some preset threshold α. Note that $\widetilde{\mathbf{C}}$ needs to be initialized to have full rank; otherwise, the inverse in (17) cannot be taken.

Assuming $\mathbf{C}$ has full row rank. Initializing $\mathbf{C}$ in the algorithm shown in Fig. 4 to be $\mathbf{C}(0)$, it is obvious that

$$\mathbf{C}(0) = \mathbf{U}_{\mathbf{C}(0)}[\boldsymbol{\Sigma}_{\mathbf{C}(0)} \quad \mathbf{0}_{N_t\times(K+L-N_t)}]\mathbf{V}_{\mathbf{C}(0)}^H \tag{15}$$

where $\mathbf{U}_{\mathbf{C}(0)}$, $\mathbf{V}_{\mathbf{C}(0)}$, and $\boldsymbol{\Sigma}_{\mathbf{C}(0)}$ are the left and singular vector matrix of $\mathbf{C}(0)$ and the invertible portion of the singular value matrix of $\mathbf{C}(0)$, respectively. Hence, to satisfy the condition that $\mathbf{C}\mathbf{P}^H = \mathbf{0}_{N_t\times K}$, it is necessary that $\mathbf{V}_{\mathbf{C}(0)} = \mathbf{U}_{\mathbf{Q}}$, where $\mathbf{U}_{\mathbf{Q}}$ is the eigenvector matrix of $\mathbf{Q}\mathbf{Q}^H$. That is

$$\begin{aligned}
\mathbf{Q}\mathbf{Q}^H &= \mathbf{U}_{\mathbf{Q}}\boldsymbol{\Lambda}_{\mathbf{Q}}\mathbf{U}_{\mathbf{Q}}^H \\
&= \mathbf{U}_{\mathbf{Q}}\begin{bmatrix}\boldsymbol{\Lambda}'_{\mathbf{Q}} & \mathbf{0}_{N_t\times(K+L-N_t)} \\ \mathbf{0}_{(K+L-N_t)\times N_t} & \mathbf{0}_{(K+L-N_t)\times(K+L-N_t)}\end{bmatrix}\mathbf{U}_{\mathbf{Q}}^H
\end{aligned}$$

where $\boldsymbol{\Lambda}'_{\mathbf{Q}} \in \mathbb{C}^{N_t\times N_t}$ is a diagonal matrix containing the nonzero eigenvalues of $\mathbf{Q}\mathbf{Q}^H$. Assume that the diagonal values of $\boldsymbol{\Lambda}_{\mathbf{Q}}$ are arranged in descending order. Hence, $\widetilde{\mathbf{C}}^2(n) = (\mathbf{U}_{\mathbf{C}(n)}\boldsymbol{\Lambda}_{\mathbf{C}(n)}\mathbf{U}_{\mathbf{C}(n)}^H)^* \otimes \mathbf{I}_{N_r}$, $n = 0, 1, \ldots, n_0$, where n and n_0 denote the iteration index and the iteration time when $\|\mathbf{E}(n)-\mathbf{E}(n-1)\|^2/\epsilon < \alpha$, respectively, and $\boldsymbol{\Lambda}_{\mathbf{C}(n)} = \boldsymbol{\Sigma}_{\mathbf{C}(n)}\boldsymbol{\Sigma}_{\mathbf{C}(n)}^*$. Thus, $\mathbf{V}_{\mathbf{C}(n)} = \mathbf{V}_{\mathbf{C}(0)} = \mathbf{U}_{\mathbf{Q}}$, $n = 0, 1, \ldots, n_0$, and the training sequence, when convergence has been reached, becomes

$$\mathbf{C}(n_0) = \mathbf{U}_{\mathbf{C}(n_0)}\left[\boldsymbol{\Sigma}_{\mathbf{C}(n_0)} \quad \mathbf{0}_{N_t\times(K+L-N_t)}\right]\mathbf{U}_{\mathbf{Q}}^H \tag{16}$$

where $\mathbf{U}_{\mathbf{C}(n_0)}$ is the singular vector matrix for $\mathbf{C}(n_0)$, and $\boldsymbol{\Sigma}_{\mathbf{C}(n_0)} \in \mathbb{C}^{N_t\times N_t}$ is a singular value matrix of $\mathbf{C}(n_0)$ containing all nonzero singular values. This conforms with the structure previously derived in [20].

Define:
- $\mathbf{C}(n)$: training sequence matrix at the n^{th} iteration
- $\mathbf{E}(n)$: error mismatch matrix at the n^{th} iteration
- $\widetilde{\mathbf{C}}(n) \triangleq (\mathbf{C}(n)\mathbf{Q})^T \otimes \mathbf{I}_{N_r}$
- $\widetilde{\mathbf{C}}^2(n) \triangleq \widetilde{\mathbf{C}}^H(n)\widetilde{\mathbf{C}}(n)$

Algorithm:
- initialize $\mathbf{C}(0)$ to any full rank matrix, s.t. $\mathbf{C}(0)\mathbf{P}^H = \mathbf{0}_{N_t\times K}$
- initialize $\mathbf{E}(1) = \mathbf{0}_{N_rN_t\times N_rN_t}$
- initialize α to the value of a given threshold
- initialize ε to the value of a given error bound
- initialize ρ to a given value for the diagonal loading
- $n = 1$

DO:
- if $rank\left(\widetilde{\mathbf{C}}^2(n)\right) < N_tN_r$
  - $\widetilde{\mathbf{C}}^2(n) = \widetilde{\mathbf{C}}^2(n) + \rho\mathbf{I}_{N_rN_t}$
- endif

$$\begin{aligned}
\mathbf{E}_{opt}(n) = \arg \min_{\substack{\|\mathbf{E}(n)\|_F\le\varepsilon \\ tr(\widehat{\mathbf{R}}+\mathbf{E}(n))=N_tN_r}} tr\Big(\Big[\widetilde{\mathbf{C}}^2(n-1) + \sigma_{\mathbf{nn}}^{-2} \\
\widetilde{\mathbf{C}}^2(n-1)\left(\widehat{\mathbf{R}}+\mathbf{E}(n)\right)\widetilde{\mathbf{C}}^2(n-1)\Big]^{-1}\Big)
\end{aligned} \tag{17}$$

$$\begin{aligned}
\widetilde{\mathbf{C}}_{opt}(n) = \arg \min_{\|\widetilde{\mathbf{C}}(n)\|_F^2\le\widetilde{P}_T} tr\Big(\Big[\Big(\widehat{\mathbf{R}}+\mathbf{E}_{opt}(n)\Big)^{-1} \\
+\sigma_{\mathbf{nn}}^{-2}\widetilde{\mathbf{C}}^H(n)\widetilde{\mathbf{C}}(n)\Big]^{-1}\Big)
\end{aligned} \tag{18}$$

- $\widetilde{\mathbf{C}}(n) = \widetilde{\mathbf{C}}_{opt}(n)$
- $n = n+1$
- $\mathbf{E}(n) = \mathbf{E}_{opt}(n)$

Until $\frac{\|\mathbf{E}(n)-\mathbf{E}(n-1)\|^2}{\varepsilon} < \alpha$

Fig. 4. Algorithm pseudocode for training sequence design.

B. Convergence Analysis

Theorem 1: The iteration depicted in Fig. 4 will always converge to the global optimal solution given that $\mathbf{C}$ is initialized as a matrix with full rank, where the constraint $\mathbf{C}(0)\mathbf{P}^H = \mathbf{0}_{N_t\times K}$ is satisfied.

Proof: Define the convex sets $\mathcal{E} = \{(\widetilde{\mathbf{C}}, \mathbf{E}) | \|\mathbf{E}\|_F \le \varepsilon\}$ and $\mathcal{C} = \{(\widetilde{\mathbf{C}}, \mathbf{E}) | \|\widetilde{\mathbf{C}}\|_F^2 \le \widetilde{P}_T\}$ containing elements that are

two tuples.[1] Express the objective function in (11) as $f(\widetilde{\mathbf{C}}, \mathbf{E})$. Given $\widetilde{\mathbf{C}}$, it is clear that $\max f(\widetilde{\mathbf{C}}, \mathbf{E})$ with respect to $\mathbf{E}$ is a nonexpansive operator, i.e.,

$$\begin{aligned} 0 &= \left\| \max_{\|\mathbf{E}_1\|_F \le \varepsilon} f(\widetilde{\mathbf{C}}, \mathbf{E}_1) - \max_{\|\mathbf{E}_2\|_F \le \varepsilon} f(\widetilde{\mathbf{C}}, \mathbf{E}_2) \right\| \\ &\le \|\mathbf{E}_1 - \mathbf{E}_2\|. \end{aligned} \tag{19}$$

Similarly, given $\mathbf{E}$, $\min f(\widetilde{\mathbf{C}}, \mathbf{E})$ with respect to $\widetilde{\mathbf{C}}$ is also a nonexpansive operator, i.e.,

$$\begin{aligned} 0 &= \left\| \min_{\|\widetilde{\mathbf{C}}_1\|_F^2 \le \widetilde{P}_T} f(\widetilde{\mathbf{C}}_1, \mathbf{E}) - \min_{\|\widetilde{\mathbf{C}}_2\|_F^2 \le \widetilde{P}_T} f(\widetilde{\mathbf{C}}_2, \mathbf{E}) \right\| \\ &\le \|\widetilde{\mathbf{C}}_1 - \widetilde{\mathbf{C}}_2\|. \end{aligned} \tag{20}$$

Moreover, the solutions for (19) and (20) will always belong to $\mathcal{E}$ and $\mathcal{C}$, respectively. Then, according to the theory of alternating projections [31], the algorithm depicted in Fig. 4 will always converge given appropriate initial conditions. Since the two sets are convex, there is a unique point of intersection, and thus, the solution obtained in Fig. 4 will always be the global optimal solution. ■

Note that it is possible for $\mathbf{C}(n)$ to lose rank when the SNR is low and when ε is sufficiently small (e.g., SNR = 0 dB and $\varepsilon = 0.1$). To prevent this from occurring, $\widetilde{\mathbf{C}}^2(n)$ is diagonally loaded, i.e., $\widetilde{\mathbf{C}}^2(n) = \widetilde{\mathbf{C}}^2(n) + \rho \mathbf{I}_{N_t N_r}$, where ρ is a small value compared with $\|\widetilde{\mathbf{C}}^2(n)\|_F$, e.g., $\rho = 0.01\|\widetilde{\mathbf{C}}^2(n)\|_F$.

C. Power Allocation

The use of the SIT sequence has an adverse effect on the recovery of the information-bearing signal at the receiver as the sequence reduces the power of the transmitted data signal. A suboptimal power allocation scheme for the training sequence was derived in [20], in which the effective SNR

$$\text{SNR}_{\text{eff}} = \frac{E\left[\|\widehat{\mathbf{H}}\mathbf{X}\|^2\right]}{E\left[\|\widetilde{\mathbf{H}}\mathbf{X} + \mathbf{N}\mathbf{Q}_D\|^2\right]} \tag{21}$$

was maximized, which consequently maximizes the probability of detecting the correct signal at the receiver. A similar power allocation scheme can similarly be derived with the spatial correlation mismatch matrix $\mathbf{E}$ taken into account. The method is suboptimal due to the fact that the numerator in (21) can be written as

$$\begin{aligned} E\left[\|\widehat{\mathbf{H}}\mathbf{X}\|^2\right] &= \text{tr}\left(E[\widehat{\mathbf{H}}\mathbf{X}\mathbf{X}^H\widehat{\mathbf{H}}^H]\right) \\ &= K\sigma_x^2 \text{tr}\left(E[\mathbf{H}\mathbf{H}^H] + E[\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H]\right) \\ &= K\sigma_x^2\left(\text{tr}(\widehat{\mathbf{R}} + \mathbf{E}) + \epsilon\right) \\ &= K\sigma_{\mathbf{xx}}^2(N_t N_r + \epsilon) \end{aligned} \tag{22}$$

where it has been assumed that $\text{tr}(E[\mathbf{H}\mathbf{H}^H]) = \text{tr}(\widehat{\mathbf{R}} + \mathbf{E}) = N_t N_r$, with $\epsilon = \text{tr}(E[\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H])$ denoting the mean square error of the channel. Note that the received SNR is defined as $\text{SNR} = -10\log_{10}\sigma_{\mathbf{nn}}^2$ under the assumption that the power of the received signal is normalized to 1. Thus, $\text{trace}(\widehat{\mathbf{R}} + \mathbf{E}) = N_t N_r$. In addition, it should be noted that there is an error in the expression for $E[\|\widehat{\mathbf{H}}\mathbf{X}\|^2]$ in [20] in which ϵ was preceded by a minus sign, even if it should be preceded by a plus sign instead, as indicated in (22). Hence, SNR_{eff} becomes

$$\text{SNR}_{\text{eff}} = \frac{\sigma_{\mathbf{xx}}^2(N_t N_r + \epsilon)}{\sigma_{\mathbf{xx}}^2\epsilon + \gamma} \tag{23}$$

where $\gamma = N_r\sigma_{\mathbf{nn}}^2(K/K + L)$. Using the property that if $\text{tr}(\mathbf{A}) > \text{tr}(\mathbf{B})$, then $\text{tr}(\mathbf{B}^{-1}) > \text{tr}(\mathbf{A}^{-1})$, given that $\mathbf{A}$ and $\mathbf{B}$ are positive definite matrices. It then follows that ϵ is upper bounded by $(N_r N_t/N_r + N_t)(\sigma_{\mathbf{nn}}^2/\sigma_{\mathbf{cc}}^2) = \beta(\sigma_{\mathbf{nn}}^2/\sigma_{\mathbf{cc}}^2)$. Substituting this upper bound into (22), the effective SNR is then lower bounded as

$$\text{SNR}_{\text{eff}}\left(\sigma_{\mathbf{cc}}^2\right) \ge \frac{\left(1 - \sigma_{\mathbf{cc}}^2\right)\left[(N_t N_r +)\sigma_{\mathbf{cc}}^2 + \beta\sigma_{\mathbf{nn}}^2\right]}{\beta\sigma_{\mathbf{nn}}^2 - \beta\sigma_{\mathbf{nn}}^2\sigma_{\mathbf{cc}}^2 + \gamma\sigma_{\mathbf{cc}}^2}. \tag{24}$$

The maximum of the effective SNR can then be achieved by maximizing the lower bound in (24), which can be accomplished by differentiating the bound with respect to $\sigma_{\mathbf{cc}}^2$, setting the result to zero, and solving for $\sigma_{\mathbf{cc}}^2$. This results in the suboptimal power allocation for the SIT sequence

$$\sigma_{\mathbf{cc},\text{subopt}}^2 = \frac{\delta\beta\sigma_{\mathbf{nn}}^2 - \sqrt{\delta\gamma\beta\sigma_{\mathbf{nn}}^2\left(\delta - \gamma + \beta\sigma_{\mathbf{nn}}^2\right)}}{\delta\left(\beta\sigma_{\mathbf{nn}}^2\right)} \tag{25}$$

where $\delta = N_t N_r$. $\sigma_{\mathbf{cc},\text{subopt}}^2$ in (25) is similar to the expression derived in [20] except for the sign error, as previously indicated. The difference is due to the sign error in (22). However, the foregoing power allocation expression is derived directly with inclusion of the spatial correlation mismatch, thus generalizing the result previously reported in [20].

IV. Simulation Results

Monte Carlo simulations were used to demonstrate the robustness of the proposed scheme. The channels used in all the simulations are assumed to be quasi-static block Rayleigh fading and spatially correlated, unless otherwise specified. The one-ring model [30] is used to generate entries of the Cholesky factors of the spatial transmit and receive correlation matrices

$$\Sigma_t(m,n) \approx J_0\left(\Delta\frac{2\pi}{\lambda}d_t|m-n|\right) \tag{26}$$

$$\Sigma_r(i,j) \approx J_0\left(\frac{2\pi}{\lambda}d_r|i-j|\right) \tag{27}$$

where d_t and d_r are the spacing between transmit and receive antennas, respectively. Δ denotes the angular spread, λ denotes the carrier wavelength, and J_0 is the 0th-order Bessel function of the first kind. The power allocation scheme in (25) for the training sequence is adopted. Quaternary phase-shift keying and Alamouti STBC are used for modulation of the information-bearing signals. In all the simulations, the threshold for the iteration algorithm is $\alpha = 10^{-6}$.

[1] Both sets are convex because their respective constraints form a norm ball.

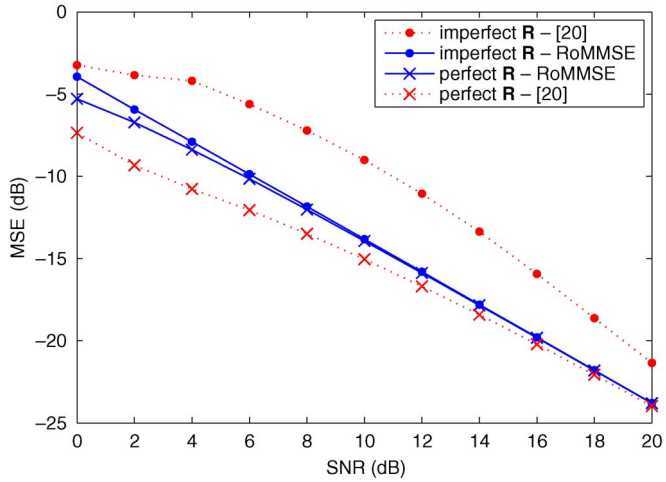

Fig. 5. MSE versus SNR performance comparison between RoMMSE and [20] for spatially correlated 2×2 MIMO channel. $\Delta = 5°$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

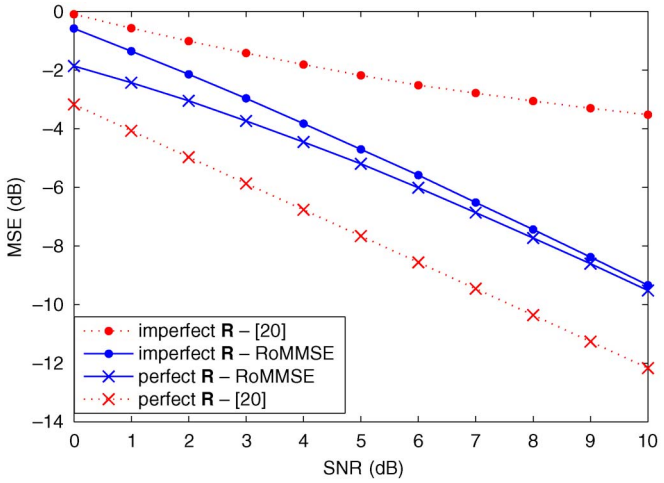

Fig. 7. MSE versus SNR performance comparison between RoMMSE and [20] for spatially correlated 4×4 MIMO channel. $\Delta = 15°$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

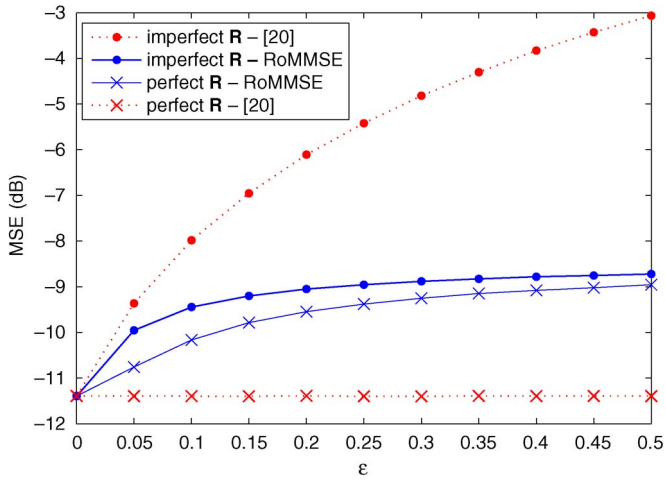

Fig. 6. MSE versus ε performance comparison between RoMMSE and [20] for spatially correlated 2×2 MIMO channel. $\Delta = 5°$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. SNR $= 5$ dB.

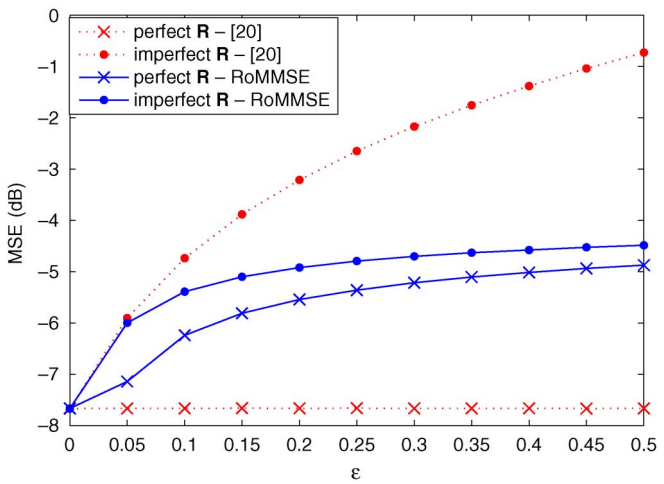

Fig. 8. MSE versus ε performance comparison between RoMMSE and [20] for spatially correlated 4×4 MIMO channel. $\Delta = 15°$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. SNR $= 5$ dB.

A 2×2 spatially correlated MIMO system with $\Delta = 5°$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$ is considered in Fig. 5. The data block size K is 60, and $L = N_t = 2$. When the correlation matrix $\mathbf{R}$ is estimated perfectly, i.e., $\mathbf{R} = \widehat{\mathbf{R}}$, the sequence design in [20] outperforms the proposed RoMMSE algorithm with $\varepsilon = 0.3$. This is the case because the sequence design in [20] is MMSE optimal when perfect knowledge of $\mathbf{R}$ is available. However, when $\mathbf{R}$ is not estimated accurately, i.e., $\mathbf{R} = \hat{\mathbf{R}} + \mathbf{E}$, the proposed RoMMSE estimator outperforms the estimator in [20] by as much as 8 dB.

Fig. 6 compares the MSE performance of the proposed scheme with that of [20] when the spatial correlation matrix error power is varied. The channel parameters in this figure are identical to those in Fig. 5 with SNR $= 5$ dB. For the case of "imperfect $\mathbf{R}$", i.e., $\mathbf{E} \neq \mathbf{0}_{N_r N_t \times N_r N_t}$, the results for [20] are obtained by solving (8) with $\mathbf{R} = \widehat{\mathbf{R}}$. In the case of "perfect $\mathbf{R}$", i.e., $\mathbf{E} = \mathbf{0}_{N_r N_t \times N_r N_t}$, the exact matrix channel correlation matrix is used to design the training sequence for both algorithms. It can be seen from the figure that the algorithm in [20] outperforms the proposed scheme when an accurate spatial correlation matrix is available for estimation. However, when $\widehat{\mathbf{R}} \neq \mathbf{R}$, then the proposed scheme outperforms [20]. Moreover, as the estimation error ε increases, the MSE of the RoMMSE estimator rises only gradually, whereas the MSE increases unbounded for [20].

Figs. 7 and 8 illustrate the same performance comparison as Figs. 5 and 6 but for 4×4 MIMO systems. The angular spread Δ is set to be 15°, and the antenna spacing d_t and d_r are 0.5λ and 0.2λ, respectively. $K = 60$, and $L = N_t = 4$. $\varepsilon = 0.3$ is used in Fig. 7, whereas SNR $= 5$ dB is used for Fig. 8. From Fig. 7, the performance of both algorithms for the 4×4 system follows the same pattern as that of the 2×2 system. Specifically, the proposed RoMMSE estimator outperforms the estimator in [20] by as much as 9 dB when MSE $= -1$ dB. In addition, unlike the algorithm in [20], the RoMMSE estimator performance does not flatten out as the SNR increases. This is because the inaccuracy in $\mathbf{R}$ has been taken into account during the channel estimation process. However, since there are more parameters to be estimated in the 4×4 system compared with the 2×2 system, there is a performance degradation not only

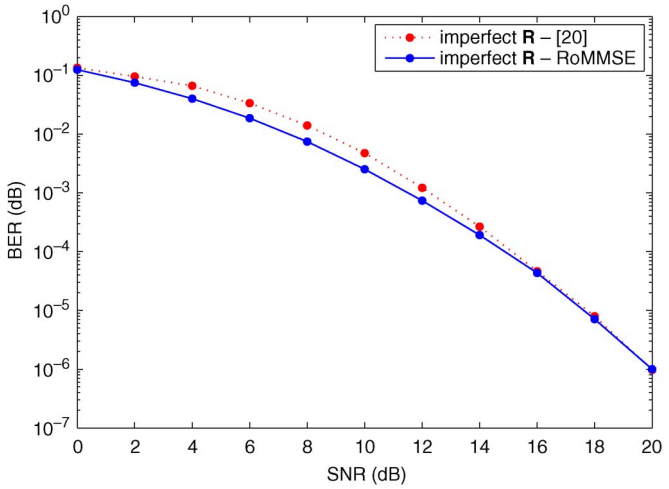

Fig. 9. BER versus SNR performance comparison between RoMMSE and [20] for spatially correlated 2×2 MIMO channel. $\Delta = 5^\circ$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

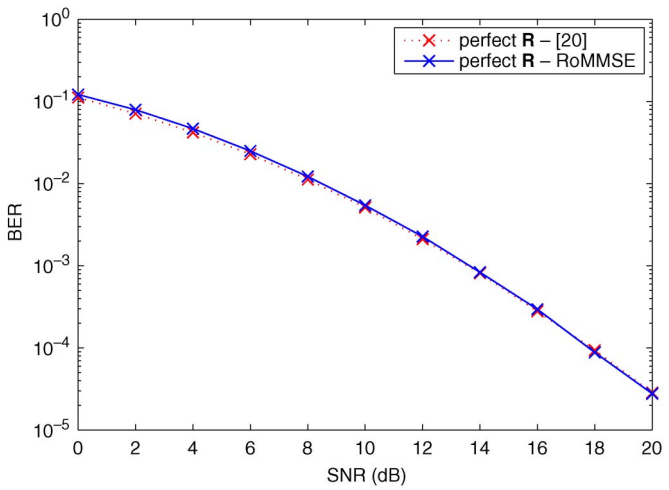

Fig. 10. BER versus SNR performance comparison between RoMMSE and [20] for spatially correlated 2×2 MIMO channel. $\Delta = 5^\circ$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

in terms of the absolute MSE, but the rate of decrease of the MSE has also diminished.

In addition to MSE performance, Figs. 9 and 10 compare the BER performance of the RoMMSE algorithm and that of [20] when the estimate of spatial correlation is imperfect and when it is perfect, respectively. The 2×2 MIMO systems are used. From Fig. 9, it can be seen that the RoMMSE algorithm outperforms the algorithm in [20] by 2 dB when the SNR is low. However, when the spatial correlation has been perfectly estimated, the RoMMSE algorithm and the algorithm in [20] render identical performance.

Figs. 11 and 12 compare the MSE performance for spatially correlated and uncorrelated MIMO channels of the RoMMSE estimator to the RMMSE and LS-RMMSE estimators in [21]. The spatial correlation in Fig. 11 is created by letting $\Delta = 5^\circ$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. The RMMSE uses diagonal loading to derive an MMSE estimator that requires only knowledge of $\mathrm{tr}(\mathbf{R})$ instead of $\mathbf{R}$ to estimate the MIMO channel. LS-RMMSE further relaxes the requirement in RMMSE by using

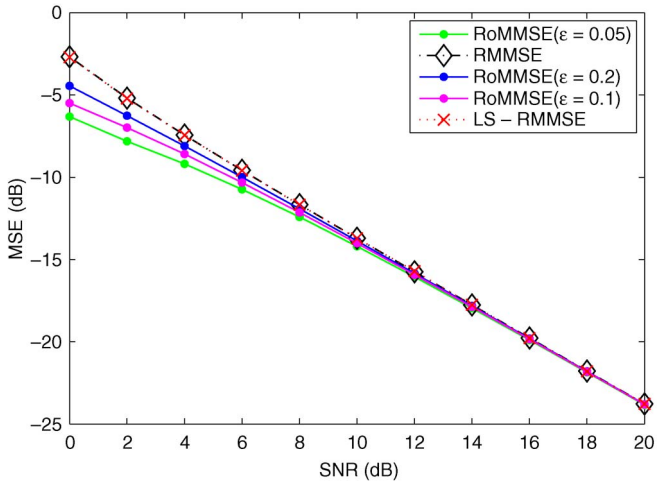

Fig. 11. MSE versus SNR performance comparison between RoMMSE, LS-RMMSE, and RMMSE [21] for spatially correlated 2×2 MIMO system. $\Delta = 5^\circ$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$.

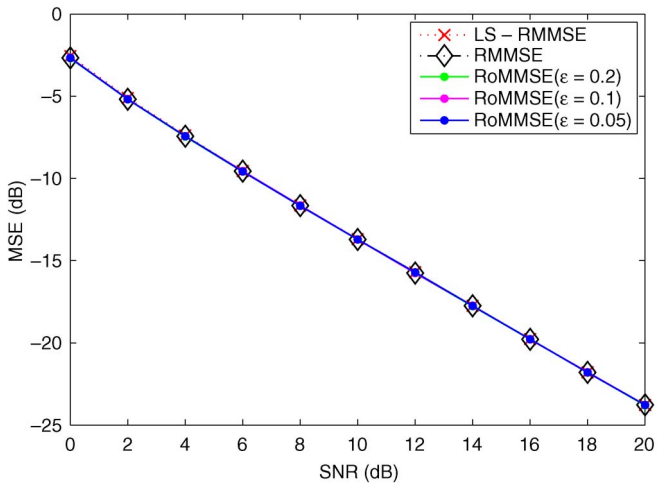

Fig. 12. MSE versus SNR performance comparison between RoMMSE, LS-RMMSE, and RMMSE [21] for spatially uncorrelated 2×2 MIMO system, i.e., $\mathbf{R} = \mathbf{I}_{N_r N_t}$. $\varepsilon = 0.3$.

TABLE I
NUMBER OF EIGENMODES USED DURING CHANNEL ESTIMATION FOR SPATIALLY CORRELATED MIMO CHANNEL

	Low SNR	High SNR
$\mathbf{E} = \mathbf{0}_{N_r N_t \times N_r N_t}$	one	all
$\mathbf{E} \neq \mathbf{0}_{N_r N_t \times N_r N_t}$	all	all

the LS method to derive an MMSE estimator that no longer requires knowledge of $\mathrm{tr}(\mathbf{R})$. Instead, only knowledge about the Frobenius norm of the received signal matrix is required. As seen in Fig. 11, when spatial correlation exists, the proposed RoMMSE algorithm outperforms the RMMSE and LS-RMMSE algorithms in low SNR by 4 dB when $\varepsilon = 0.05$ but only by about 2 dB when $\epsilon = 0.2$. This shows that the error power bound cannot be too high; otherwise, the performance of the proposed scheme will degrade. This is so because as ε increases, $\mathbf{E}$ obtained from the iterative algorithm will decorrelate the spatial correlation more, thus adversely affecting the performance of the proposed scheme. This can be explained as follows. The RoMMSE estimator strives to minimize the

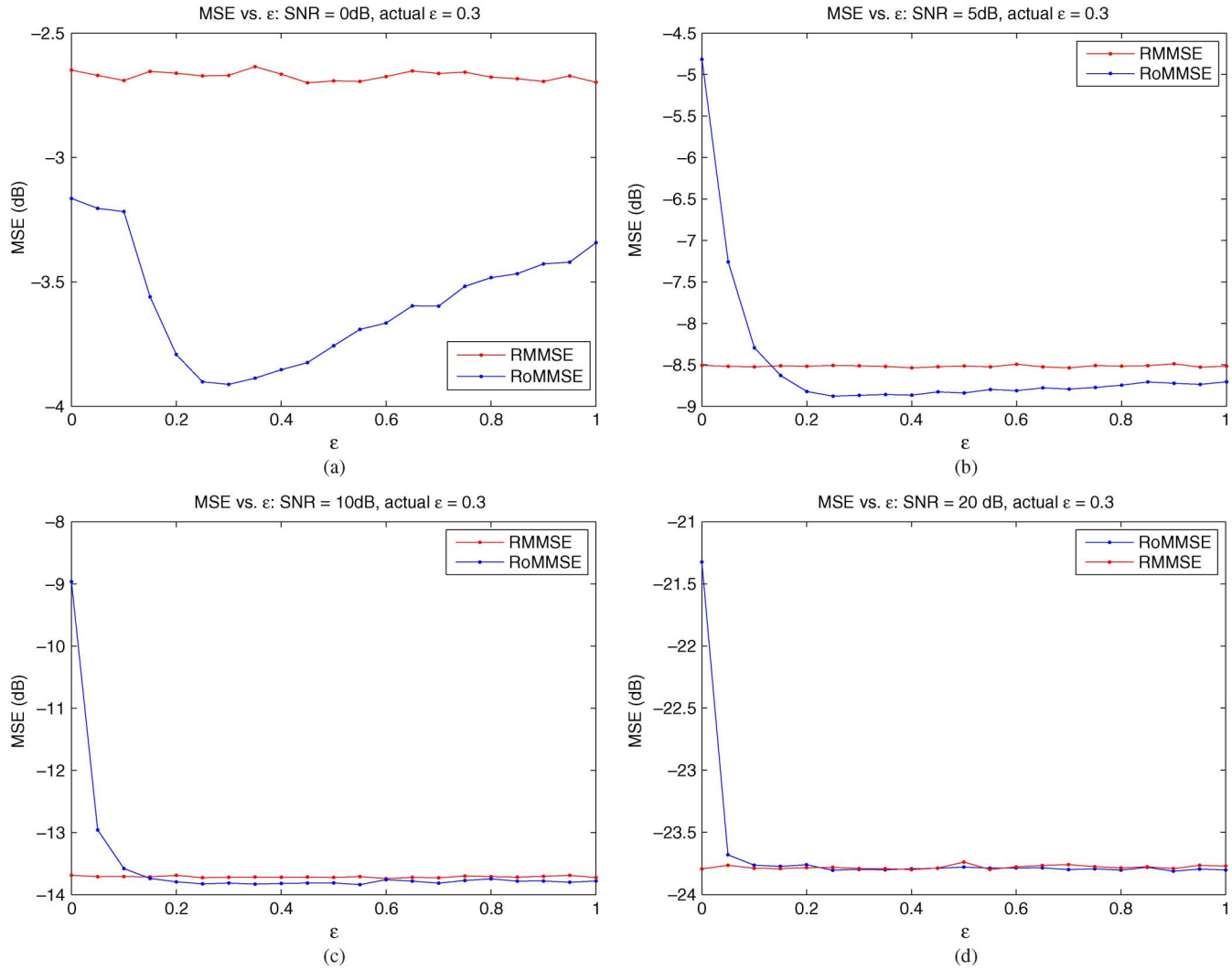

Fig. 13. MSE versus ε performance comparison between RoMMSE and RMMSE [21] for spatially correlated 2×2 MIMO system. $\Delta = 5^\circ$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. Actual correlation mismatch power is 0.3. (a) SNR $= 0$ dB. (b) SNR $= 5$ dB. (c) SNR $= 10$ dB. (d) SNR $= 20$ dB.

worst case MSE, as seen in (11). The worst-case MSE can be attained by increasing the number of parameters that needs to be estimated, the maximum being $N_r N_t$. In other words, the present method attempts to increase the degrees of freedom in the correlated MIMO channel by reducing the spatial correlation. As ε increases, $\|\mathbf{E}\|_F$ also increases, which allows $\mathbf{E}$ more freedom to zero out the off-diagonal elements of $\mathbf{R}$, therefore lessening the spatial correlation. Complete decorrelation of $\mathbf{R}$ is attained as $\varepsilon \to \infty$. In addition, it has been observed that when the threshold α is met, $\mathbf{E}$ and $\widetilde{\mathbf{C}}^2$ share the same eigenvector matrix as $\widehat{\mathbf{R}}$. This has been proven analytically in Appendix A. When there is no spatial correlation mismatch, it was shown in [12] that the transmitted signal corresponds to transmitting in specific eigenmodes of the spatial correlation, which determines which particular eigenmode of the channel will be estimated. Furthermore, the power on each eigenmode is determined by waterfilling solution based on certain optimization criteria, such as minimum MSE and maximum conditional mutual information. When the SNR is low, it was found that all the power will be allocated to the strongest eigenmode. However, when the SNR is high, the power is evenly distributed among all the eigenmodes. When spatial correlation mismatch has been accounted for, it can be seen from the simulations that regardless of whether the system is operating under low or high SNR, the mismatch matrix $\mathbf{E}$ not only decorrelates the channel but equalizes all the diagonal values of $\widehat{\mathbf{R}}$ as well, such that $\mathrm{tr}(\widehat{\mathbf{R}} + \mathbf{E}) = N_r N_t$ given that ε is sufficiently large. Note that this is also true even when $\mathrm{tr}(\widehat{\mathbf{R}} + \mathbf{E}) = N_r N_t$ is not a constraint in (17). Hence, the robust training sequence evenly distributes power across all the eigenmodes. This is because the worst-case mismatch matrix $\mathbf{E}_w$ can be obtained only when $\widehat{\mathbf{R}}$, $\mathbf{E}$, and $\widetilde{\mathbf{C}}^2$ are all diagonalized and because $\mathbf{E}$ and $\widetilde{\mathbf{C}}^2$ share the same eigenvectors as $\widehat{\mathbf{R}}$ (see Appendix A). The constraint $\mathrm{tr}(\widehat{\mathbf{R}} + \mathbf{E}) = N_r N_t$ forces $\mathbf{E}$ to diagonalize $\widehat{\mathbf{R}}$ and equalizes the diagonal values of $\widehat{\mathbf{R}}$ such that the constraint is satisfied. Hence, the mismatch matrix $\mathbf{E}$ will evenly distribute power across all the eigenmodes of $\mathbf{R}$. If ε is not sufficient large, then $\mathbf{E}$ will not have enough degrees of freedom to diagonalize and equalize the diagonal values of $\widehat{\mathbf{R}}$. This phenomenon has been summarized in Table I.

The MSE performance for different values of ε is compared between RoMMSE and RMMSE when the actual mismatch error power is equal to 0.3. The results in Fig. 13 suggest that ε should not be chosen to be smaller than the actual mismatch

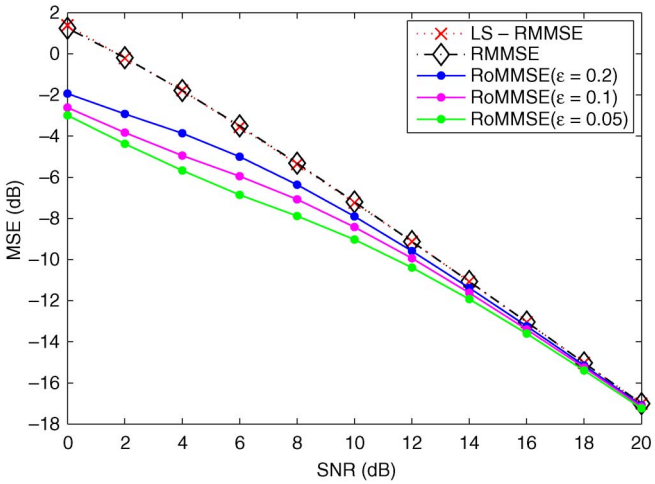

Fig. 14. MSE versus SNR performance comparison using time-multiplexed pilots between RoMMSE, LS-RMMSE, and RMMSE [21] for spatially correlated 2×2 MIMO system. $\Delta = 5^\circ$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$.

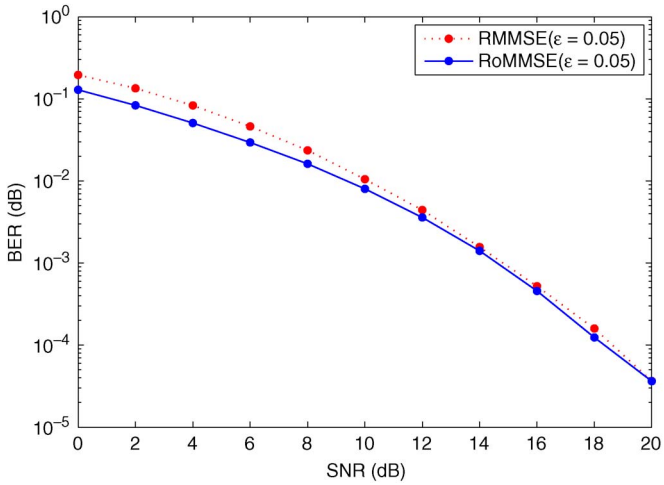

Fig. 15. BER versus SNR performance comparison using time-multiplexed pilots between RoMMSE and RMMSE [21] for spatially correlated 2×2 MIMO channel. $\Delta = 5^\circ$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. $\varepsilon = 0.05$.

error power (if that is estimated *a priori*) as that will greatly and adversely affect the MSE. ε should not be chosen to be too high either, but that will be less detrimental as having a small ε, particularly under low SNR condition. Hence, in the absence of knowledge about the actual mismatch error power, a high value should be selected for ε. The results for higher SNR [see Fig. 13(b)–(d)] corroborate with those in Fig. 11, i.e., the MSE of the RMMSE and RoMMSE estimator converges as the SNR increases.

The performance of the RoMMSE estimator is also compared with that of RMMSE and LS-RMMSE when the MIMO channel is spatially uncorrelated. Fig. 12 indicates that in this situation all three estimators render similar MSE performance, which suggests that all three estimators are identical. This is indeed the case, and it is proven in Appendix B.

The data detection performance for the RMMSE and RoMMSE algorithms is also compared in the case of spatially correlated and uncorrelated channels. Since the RMMSE algorithm is proposed in a time-multiplexed pilot scheme, both MSE and BER are compared using time-multiplexed pilots. The channel estimation performance using time-multiplexed pilots is shown in Fig. 14, where the RoMMSE algorithm outperforms the RMMSE and LS-RMMSE algorithms, similar to the performance shown in Fig. 11. Notice that the estimation performance of the RoMMSE algorithm in Fig. 14 is worse than that shown in Fig. 11. This is because the power of the pilot is less than that of the SIT sequence. Next, the BER performance comparison is shown in Fig. 15. With $\varepsilon = 0.05$, RoMMSE outperforms the RMMSE by 2.5 dB in the low SNR region.

The BER performance of the RoMMSE algorithm versus different values of ε is shown in Fig. 16. Notice that a lower BER is obtained with increasing ε. As previously explained, this is because the mismatch matrix has more freedom to decorrelate the spatial correlation matrix as ε increases, which enhances the spatial diversity of the system, thus improving the BER performance.

Table II shows the number of iterations needed before the algorithm in Fig. 4 converges under different initial conditions

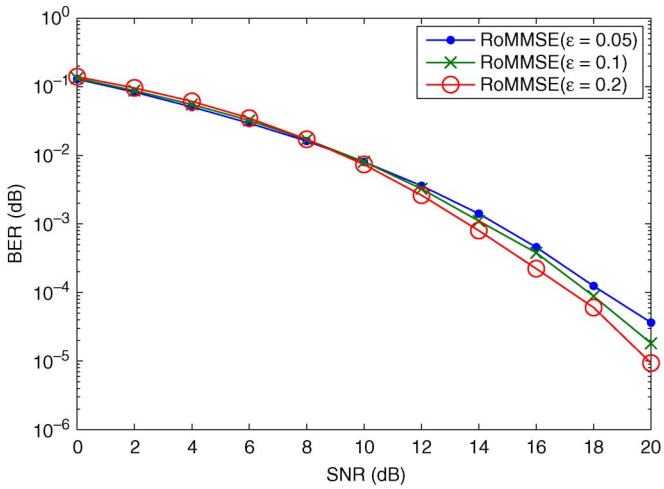

Fig. 16. BER versus SNR performance comparison using time-multiplexed pilots of RoMMSE with different ϵ for spatially correlated 2×2 MIMO channel. $\Delta = 5^\circ$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$.

TABLE II
AVERAGE NUMBER OF ITERATIONS REQUIRED FOR CONVERGENCE FOR THE PROPOSED ROMMSE ESTIMATOR. $d_t = 0.5\lambda$, $d_r = 0.2\lambda$, $\varepsilon = 0.3$, AND SNR = 5 dB, ANGULAR SPREAD = 15° (4×4), 5° (2×2)

$N_t \times N_r$	$\mathbf{C}(0)$	Average number of iterations
4×4	randomly generated	5.742
4×4	orthogonal	4.000
2×2	randomly generated	3.524
2×2	orthogonal	3.000

$\mathbf{C}(0)$. It shows that the proposed iterative algorithm always, on the average, converges faster if $\mathbf{C}(0)$ is initialized to be an orthogonal matrix than when it is initialized to be a random matrix. Although the table only shows the performance when the SNR is 5 dB, this convergence behavior has been observed for all the SNR values that have been tested. This speed up is due to the fact that the worst case MSE is achieved if the argument inside the trace operator in (17) forms a diagonal matrix. Hence, if $\mathbf{C}(0)$ is initialized to be an orthogonal matrix, then $\widetilde{\mathbf{C}}(n)$, $n > 0$ will be closer to the optimal solution than when $\mathbf{C}(0)$ is initialized to be a random matrix, as it is already

TABLE III
AVERAGE NUMBER OF ITERATIONS REQUIRED FOR THE IMPLEMENTATION OF THE SDPT3 ALGORITHM FOR ROMMSE AND [20]. $d_t = 0.5\lambda$, $d_r = 0.2\lambda$, $\varepsilon = 0.3$, ANGULAR SPREAD $= 15^\circ$ (4×4), AND 5° (2×2). ORTHOGONAL MATRIX IS USED TO INITIATE $\mathbf{E}(0)$ IN ROMMSE

$N_t \times N_r$	RoMMSE	[20]
4×4	303.5	15.5
2×2	56.5	11.17

equal to a diagonal matrix of the form $\alpha \mathbf{I}_{N_r N_t}$, for α being equal to an arbitrary constant.

Table III shows the average number of iterations required for the proposed RoMMSE and the algorithm in [20]. Note that the computational complexity in [12] is similar to that of [20]. The complexity figures depicted in Table III are the average number of iterations used in the primal–dual interior-point method in SDPT3 [33]. This is the default solver used in *cvx*, and it is also used to find solutions for (17) and (18). The entries in Table III are computed by noting the number of times (17) and (18) are invoked, multiplied by the number of iterations required to obtain optimal values in (17) and (18). The values obtained for the low and high SNR conditions are then averaged to obtain the entries in the table. The computational complexity for RoMMSE is understandably higher than those of [20] and [12] because the RoMMSE algorithm requires an extra step to solve for the worst-case $\mathbf{E}$ to compensate for the spatial correlation mismatch.

V. CONCLUSION

A robust SIT sequence design algorithm for spatially correlated MIMO channel estimation has been proposed. The algorithm has been shown to be robust against error in the spatial correlation estimate. When the robust training sequence is inserted into the MMSE estimator, a robust MMSE, or RoMMSE, estimator is derived. Simulation results have shown that the proposed RoMMSE estimator not only outperforms the optimal MMSE estimator in [20] when the error in the spatial correlation exists, albeit at the cost of higher computational complexity, but it also outperforms other robust designs, such as RMMSE and LS-RMMSE [21]. Furthermore, it has been shown that the spatial correlation mismatch matrix decorrelates the spatial correlation matrix when the error power bound goes to infinity. Finally, the RoMMSE scheme is identical to the RMMSE-based schemes when the MIMO channel is spatially uncorrelated.

APPENDIX A
DECORRELATION OF $\widehat{\mathbf{R}}$

Insert (14) into (11), and note that the SVDs of $\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}$ and $\widehat{\mathbf{R}}$ are $\mathbf{V}_{\widetilde{\mathbf{C}}^2}\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}\mathbf{V}_{\widetilde{\mathbf{C}}^2}^H$ and $\mathbf{U}_{\widehat{\mathbf{R}}}\mathbf{\Lambda}_{\widehat{\mathbf{R}}}\mathbf{U}_{\widehat{\mathbf{R}}}^H$, respectively. The sequence design problem in (11) is equivalent to solving (17) and (18) iteratively. Using the SVD, (17) can be rewritten as

$$\min_{\|\mathbf{E}\|_F \le \varepsilon} \operatorname{tr}\left(\left[\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} \mathbf{V}_{\widetilde{\mathbf{C}}^2}^H \left(\mathbf{U}_{\widehat{\mathbf{R}}} \mathbf{\Lambda}_{\widehat{\mathbf{R}}} \mathbf{U}_{\widehat{\mathbf{R}}}^H + \mathbf{E} \right) \mathbf{V}_{\widetilde{\mathbf{C}}^2} \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} \right]^{-1} \right).$$

Define $\mathbf{A} \triangleq \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} \mathbf{V}_{\widetilde{\mathbf{C}}^2}^H (\mathbf{U}_{\widehat{\mathbf{R}}} \mathbf{\Lambda}_{\widehat{\mathbf{R}}} \mathbf{U}_{\widehat{\mathbf{R}}}^H + \mathbf{E}) \mathbf{V}_{\widetilde{\mathbf{C}}^2} \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}$. The objective function can then be written as $f(\boldsymbol{\lambda}(\mathbf{A})) = \sum_i (1/\lambda_i(\mathbf{A}))$, where $\boldsymbol{\lambda}(\mathbf{A})$ denotes a vector composed of eigenvalues of $\mathbf{A}$. Since $\phi(\lambda_i(\mathbf{A})) = 1/\lambda_i(\mathbf{A})$ is a convex function, $f(\boldsymbol{\lambda}(\mathbf{A}))$ is Schur convex [12]. Moreover, $\mathbf{A}$ is a symmetric matrix. Therefore, $f(\boldsymbol{\lambda}(\mathbf{A}))$ majorizes $f(\mathbf{d}(\mathbf{A}))$, i.e., $f(\boldsymbol{\lambda}(\mathbf{A})) \ge f(\mathbf{d}(\mathbf{A}))$ [12], where $\mathbf{d}(\mathbf{A})$ denotes a vector that is composed of diagonal values of $\mathbf{A}$. Since the equality will hold when $\mathbf{A}$ is a diagonal matrix, then the worst-case mismatch error will be $\mathbf{E}_w = \mathbf{V}_{\widetilde{\mathbf{C}}} \mathbf{\Lambda}_{\mathbf{E}} \mathbf{V}_{\widetilde{\mathbf{C}}}^H$, which ensures that the lower bound of the MSE is reached. This implies that $\mathbf{V}_{\widetilde{\mathbf{C}}^2}^H \mathbf{U}_{\widehat{\mathbf{R}}}$ has to be a diagonal matrix and that $\mathbf{E}$ shares the same eigenvectors as $\widehat{\mathbf{R}}$. The first condition can thus be achieved if $\widehat{\mathbf{R}}$ and $\widetilde{\mathbf{C}}^2$ also share the same eigenvector matrix.

APPENDIX B
COMPARISON OF ROMMSE AND RELAXED MINIMUM MEAN SQUARE ERROR ESTIMATORS

When the MIMO system is spatially uncorrelated, i.e., $\mathbf{R} = \mathbf{I}_{N_r N_t}$, the channel estimate from the RMMSE channel estimator in [21] becomes

$$\begin{aligned} \widehat{\mathbf{h}}_{\mathrm{RMMSE}} &= \widetilde{\mathbf{C}}^H \left(\widetilde{\mathbf{C}} \widetilde{\mathbf{C}}^H + \frac{\sigma_{\mathbf{nn}}^2 N_r N_t}{tr(\widehat{\mathbf{R}})} \mathbf{I}_{N_r N_t} \right)^{-1} \mathbf{y} \\ &= \widetilde{\mathbf{C}}^H \left(\widetilde{\mathbf{C}} \widetilde{\mathbf{C}}^H + \sigma_{\mathbf{nn}}^2 \mathbf{I}_{N_r N_t} \right)^{-1} \mathbf{y}. \end{aligned}$$

It is assumed that orthogonal sequences are employed for the RMMSE channel estimator [21], i.e., $\mathbf{C}^H\mathbf{C} = (P_T/N_t)\mathbf{I}_{K+L}$. The estimate of the RoMMSE estimator is written as

$$\begin{aligned} &\widehat{\mathbf{h}}_{\mathrm{RoMMSE}} \\ &= (\widehat{\mathbf{R}} + \mathbf{E}_w) \widetilde{\mathbf{C}}^H \left(\widetilde{\mathbf{C}} (\widehat{\mathbf{R}} + \mathbf{E}_w) \widetilde{\mathbf{C}}^H + \sigma_{\mathbf{nn}}^2 \mathbf{I}_{N_r N_t} \right)^{-1} \mathbf{y} \\ &= (\mathbf{I}_{N_r N_t} + \mathbf{E}_w) \widetilde{\mathbf{C}}^H \left(\widetilde{\mathbf{C}} (\mathbf{I}_{N_r N_t} + \mathbf{E}_w) \widetilde{\mathbf{C}}^H + \sigma_{\mathbf{nn}}^2 \mathbf{I}_{N_r N_t} \right)^{-1} \mathbf{y} \end{aligned}$$

where $\mathbf{E}_w$ is the worst-case error of the estimated spatial correlation. Let $\mathbf{V}_{\widetilde{\mathbf{C}}^2}$ and $\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}$ denote the eigenvector and eigenvalue matrix of $\widetilde{\mathbf{C}}^2$, respectively. From (13), (17) is equivalent to

$$\begin{aligned} &\min_{\|\mathbf{E}\|_F \le \varepsilon} \operatorname{tr}\left(\left[\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} \mathbf{V}_{\widetilde{\mathbf{C}}^2}^H (\mathbf{I}_{N_r N_t} + \mathbf{E}) \mathbf{V}_{\widetilde{\mathbf{C}}^2} \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} \right]^{-1} \right) \\ &= \min_{\|\mathbf{E}\|_F \le \varepsilon} \operatorname{tr}\left(\left[\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}^2 + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} \mathbf{V}_{\widetilde{\mathbf{C}}^2}^H \mathbf{E} \mathbf{V}_{\widetilde{\mathbf{C}}^2} \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} \right]^{-1} \right). \end{aligned} \tag{28}$$

Define $\mathbf{A} \triangleq \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}^2 + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} \mathbf{V}_{\widetilde{\mathbf{C}}^2}^H \mathbf{E} \mathbf{V}_{\widetilde{\mathbf{C}}^2} \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}$. The objective function can then be written as $f(\boldsymbol{\lambda}(\mathbf{A})) = \sum_i (1/\lambda_i(\mathbf{A}))$, where $\boldsymbol{\lambda}(\mathbf{A})$ denotes a vector composed of eigenvalues of $\mathbf{A}$. Since $\phi(\lambda_i(\mathbf{A})) = 1/\lambda_i(\mathbf{A})$ is a convex function, $f(\boldsymbol{\lambda}(\mathbf{A}))$ is Schur convex [12]. Moreover, $\mathbf{A}$ is a symmetric matrix. Therefore, $f(\boldsymbol{\lambda}(\mathbf{A}))$ majorizes $f(\mathbf{d}(\mathbf{A}))$, i.e., $f(\boldsymbol{\lambda}(\mathbf{A})) \ge f(\mathbf{d}(\mathbf{A}))$ [12], where $\mathbf{d}(\mathbf{A})$ denotes a vector that is composed of diagonal values of $\mathbf{A}$. Since the equality will hold when $\mathbf{A}$ is

a diagonal matrix, then the worst-case mismatch error will be $\mathbf{E}_w = \mathbf{V}_{\widetilde{\mathbf{C}}} \mathbf{\Lambda}_{\mathbf{E}} \mathbf{V}_{\widetilde{\mathbf{C}}}^H$, which ensures that the lower bound of the MSE is reached. Since $\widehat{\mathbf{R}} = \mathbf{I}_{N_r N_t}$, from Section IV, $\mathbf{E}_w$ will have to be a diagonal matrix (or a linear combination of one) so that it will not minimize the degrees of freedom in the MIMO channel. This implies that either $\mathbf{V}_{\widetilde{\mathbf{C}}^2}$ or $\mathbf{\Lambda}_{\mathbf{E}}$ is an identity matrix. However, since $\widetilde{\mathbf{C}}^2$ is not necessarily a diagonal matrix, it is not necessary for $\mathbf{V}_{\widetilde{\mathbf{C}}^2}$ to be an identity matrix. This implies that $\mathbf{\Lambda}_{\mathbf{E}}$ must be either an identity or an all zero matrix. Since the constraint $\mathrm{tr}(\widehat{\mathbf{R}} + \mathbf{E}) = N_t N_r$ must be satisfied, therefore, $\mathbf{E}_w$ must be an all zero matrix, i.e., $\mathbf{E}_w = \mathbf{0}_{N_r N_t \times N_r N_t}$. Substituting $\mathbf{E}_w = \mathbf{0}_{N_r N_t \times N_r N_t}$ into (11) and solving for $\mathbf{C}$, $\mathbf{C}$ becomes an orthogonal matrix. That is, the optimal SIT sequence is an orthogonal sequence, which agrees with the conclusion in [21] that the optimal training sequence for spatially uncorrelated MIMO channel is an orthogonal sequence. Therefore, $\widehat{\mathbf{h}}_{\mathrm{RoMMSE}}$ becomes

$$\widehat{\mathbf{h}}_{\mathrm{RoMMSE}} = \widetilde{\mathbf{C}}^H \left(\widetilde{\mathbf{C}} \widetilde{\mathbf{C}}^H + \sigma_{\mathbf{nn}}^2 \mathbf{I}_{N_r N_t} \right)^{-1} \mathbf{y}$$

which implies that the estimation performance of the proposed RoMMSE estimator and the RMMSE estimator is identical when the MIMO channel is spatially uncorrelated, thus agreeing with the simulation results in Section IV.

REFERENCES

[1] G. J. Foschini, "Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, Summer 1996.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wirel. Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[4] C. Budianu and L. Tong, "Channel estimation for space–time orthogonal block codes," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2515–2528, Oct. 2002.

[5] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 2515–2528, Apr. 2003.

[6] B. L. Hughes, "Differential space–time modulation," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.

[7] B. M. Hochwald and T. L. Marzetta, "Unitary space–time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[8] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1169–1174, Jul. 2000.

[9] H. Jafarkhani and V. Tarokh, "Multiple transmit antenna differential detection from generalized orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2626–2631, Sep. 2001.

[10] M. Enescu, T. Roman, and V. Koivunen, "Channel estimation and tracking in spatially correlated MIMO OFDM systems," in *Proc. IEEE Workshop Statist. Signal Process.*, Sep. 2003, pp. 347–350.

[11] M. Kiessling, J. Speidel, and Y. Chen, "MIMO channel estimation in correlated fading environments," in *Proc. IEEE Veh. Technol. Conf.-Fall*, Oct. 2003, vol. 2, pp. 1187–1191.

[12] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[13] N. Czink, G. Matz, D. Seethaler, and F. Hlawatsch, "Improved MMSE estimation of correlated MIMO channels using a structured correlation estimator," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2005, pp. 595–599.

[14] B. Farhang-Boroujeny, "Pilot-based channel identification: Proposal for semi-blind identification of communication channels," *IEEE Electron. Lett.*, vol. 31, no. 13, pp. 1044–1046, Jun. 1995.

[15] G. T. Zhou, M. Viberg, and T. McKelvey, "Superimposed periodic pilots for blind channel estimation," in *Proc. 35th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2001, vol. 1, pp. 653–657.

[16] J. K. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Commun. Lett.*, vol. 7, no. 9, pp. 413–415, Sep. 2003.

[17] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel estimation using implicit training," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 240–254, Jan. 2004.

[18] D. H. Pham and J. H. Manton, "Orthogonal superimposed training on linear precoding: A new affine precoder design," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2005, pp. 445–449.

[19] A. Vosoughi and A. Scaglione, "Everything you always wanted to know about training: Guidelines derived using affine precoding framework and the CRB," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 940–954, Mar. 2006.

[20] V. Nguyen, H. D. Tuan, H. H. Nguyen, and N. N. Tran, "Optimal superimposed training design for spatially correlated fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3206–3217, Aug. 2008.

[21] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[22] Y.-C. Chen and Y. T. Su, "MIMO channel estimation in spatially correlated environments," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2004, vol. 1, pp. 498–502.

[23] B. Vucetic and J. Yuan, *Space–time Coding*. New York: Wiley, 2003.

[24] N.N. Tran, D. H. Pham, H. D. Tuan, and H. H. Nguyen, "Orthogonal affine precoding and decoding for channel estimation and source detection in MIMO Frequency-Selective Fading Channels," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1151–1162, Mar. 2009.

[25] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2113–2123, Dec. 2002.

[26] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453–466, Mar. 2005.

[27] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, 1993.

[28] H. Miao and M. J. Juntti, "Space–time channel estimation and performance analysis for wireless MIMO-OFDM systems with spatial correlation," *IEEE Trans. Veh. Technol.*, vol. 54, no. 6, pp. 2003–2016, Nov. 2005.

[29] M. Grant and S. Boyd, cvx User's Guide, Feb. 2009.

[30] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2002.

[31] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1—Theory," *IEEE Trans. Med. Imag.*, vol. MI-1, no. 2, pp. 81–94, Oct. 1982.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[33] K. C. Toh, R. H. Tütüncü, and M. J. Tood, On the Implementation and Usage of SDPT3—A MATLAB Software Package for Semidefinite-Quadratic-Linear Programming, Ver. 4.0, Jul. 2006. [Online]. Available: http://www.math.nus.edu.sg/~mattohkc/sdpt3.html

Chin-Te Chiang received the B.S. and M.S. degrees in electronics engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 2007 and 2010, respectively.

He is currently with the Intellectual Property Office, Ministry of Economic Affairs, Taipei, Taiwan. His research interests include the signal processing aspects of wireless communication systems and biomedical engineering.

Carrson C. Fung (M'91) received the B.S. degree in electrical engineering from Carnegie Mellon University, Pittsburgh, PA, in 1994, the M.S. degree in electrical engineering from Columbia University, New York, NY, in 1996, and Ph.D. degree in electrical engineering from The Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2005.

He was the recipient of the prestigious Sir Edward Youde Ph.D. Fellowship in 2001–2002. From 1994 to 1999, he was a Member of Technical Staff with AT&T and Lucent Technologies Bell Laboratories, Holmdel, NJ, where he worked on video and audio coding. He was also a Researcher with the Hong Kong Applied Science and Technology Research Institute in 2005, where he worked on multiple-input–multiple-output orthogonal frequency-division multiplexing systems, and a Senior DSP Engineer with Sennheiser Research Lab, Palo Alto, CA, in 2006, where he worked on microphone and microphone array technologies. Since 2006, he has been an Assistant Professor with the National Chiao Tung University, Hsinchu, Taiwan. His research interests include digital signal processing, communications, and biomedical engineering.